

①

The bus bandwidth and latency between the CPU and the GPU may become a bottleneck, which may be alleviated in the future by introducing interconnects with higher bandwidth. The main companions between these two interconnect technologies are:

CPU Latency memory is low whereas GPU has high bandwidth memory

CPU is random accesses. GPU is sequential accesses

CPU has 20 GB/s bandwidth. GPU has 100 GB/s bandwidth

CPU is well known programming model whereas GPU is niche programming model

The actual bandwidth and latency values will change as the technology evolves